Title	Computer Supported Content Analysis: Challenges, research and developments
Author(s)	Huang, R; Li, Y
Citation	Public Seminar on Data Mining in Education - Content & Interaction Analysis of CSCL Discourse Data for Assessing Knowledge Building Outcomes, Hong Kong, China, 27 October 2006
Issued Date	2006
URL	http://hdl.handle.net/10722/43985
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Computer Supported Content Analysis

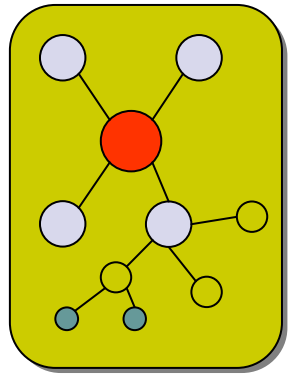
——Challenges, research and developments

Ronghuai Huang & Yanyan Li

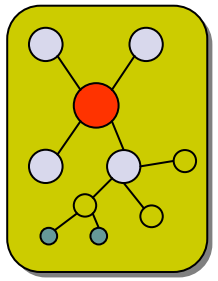
Beijing Normal University

Email: huangrh@bnu.edu.cn

Http://ksei.bnu.edu.cn

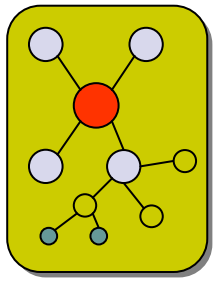


Why CSCL?



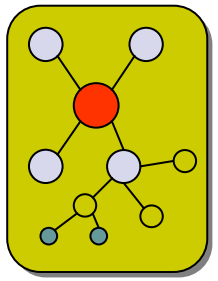
- Three issues about cooperation
  - defined as “acting together, in a coordinated way at work, or in social relationships, in the pursuit of shared goals…”
  - Is seen as central to our everyday lives
  - Cooperative learning is process driven
- A human group is a collection of individuals, who have interdependent relations, and who perceive themselves as a group that is recognised by non members
- People working cooperatively in CSCL environments do work in groups in complex ways

Why CSCL?



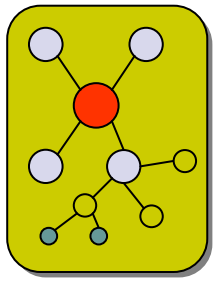
- What are the outcomes of cooperative learning?
  - Cooperative methods lead to higher achievement than competitive or individualistic ones.
  - Cooperative learning increase the positive affect of classrooms and students working cooperatively become more cooperative; they learn pro-social behaviours such as hot to get with others, how to listen and so on.
  - Cooperative learning fosters knowledge about the learning process.

Possibilities vs. Wicked problems



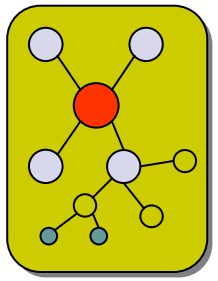
- Possibilities: Collaborative technologies are shown to enhance student motivation, self-reflection, working with complex problems, and promote collaboration between learners.
- 'wicked problems': a "problem that can be characterized as an evolving set of interlocking issues and constraints in a constantly changing context".
- Basic problems of technical infrastructure and shortage of IT-trained staff.

School culture challenges of CSCL



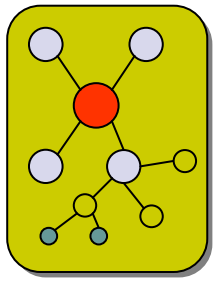
- issues concerning the compatibility of CSCL with the curriculum and the organizational structure of the school (e.g. Cullen)
- many learners seem to have great difficulties in participating in collaborative inquiry activities if these are not highly structured and if they are not given clear instructions (e.g. Blake & Rapanotti; Ploetzner et al.).
- learners often do not reach a higher level of discussion and knowledge building (e.g. Lipponen et al.; Muukkonen et al.; Mäkitalo et al.).
- many learners seem to operate under the assumption that a knowledge building process in school environment is a kind of a "question-answer-game" (see Kynigos, Dimaraki & Trouki).
- In addition, teachers were shown to have difficulties in guiding a collaborative inquiry process (e.g. Rahikainen et al.).

Pedagogical challenges of CSCL



- We do not yet seem to fully understand how technology should be employed in order to best support collaborative learning and higher-level knowledge building in different educational settings. The challenges include
  - a need for deeper knowledge about the kinds of activities that should accompany CSCL.
  - seek for an understanding of the best combination of CSCL and traditional, more individualistic instructional approaches (e.g. Muukkonen et al.).
  - unequal participation of the learners in computer supported collaborative learning. Some studies pointed out a tendencial exclusion of weaker or less motivated learners from computer-mediated discussions (e.g. Cullen, Lipponen et al.; Rahikainen et al.; Tapola et al.).
  - better understanding the kind of pedagogical support needed during computer supported collaborative learning (e.g. Salovaara & Järvelä; Tholander).

Solutions?

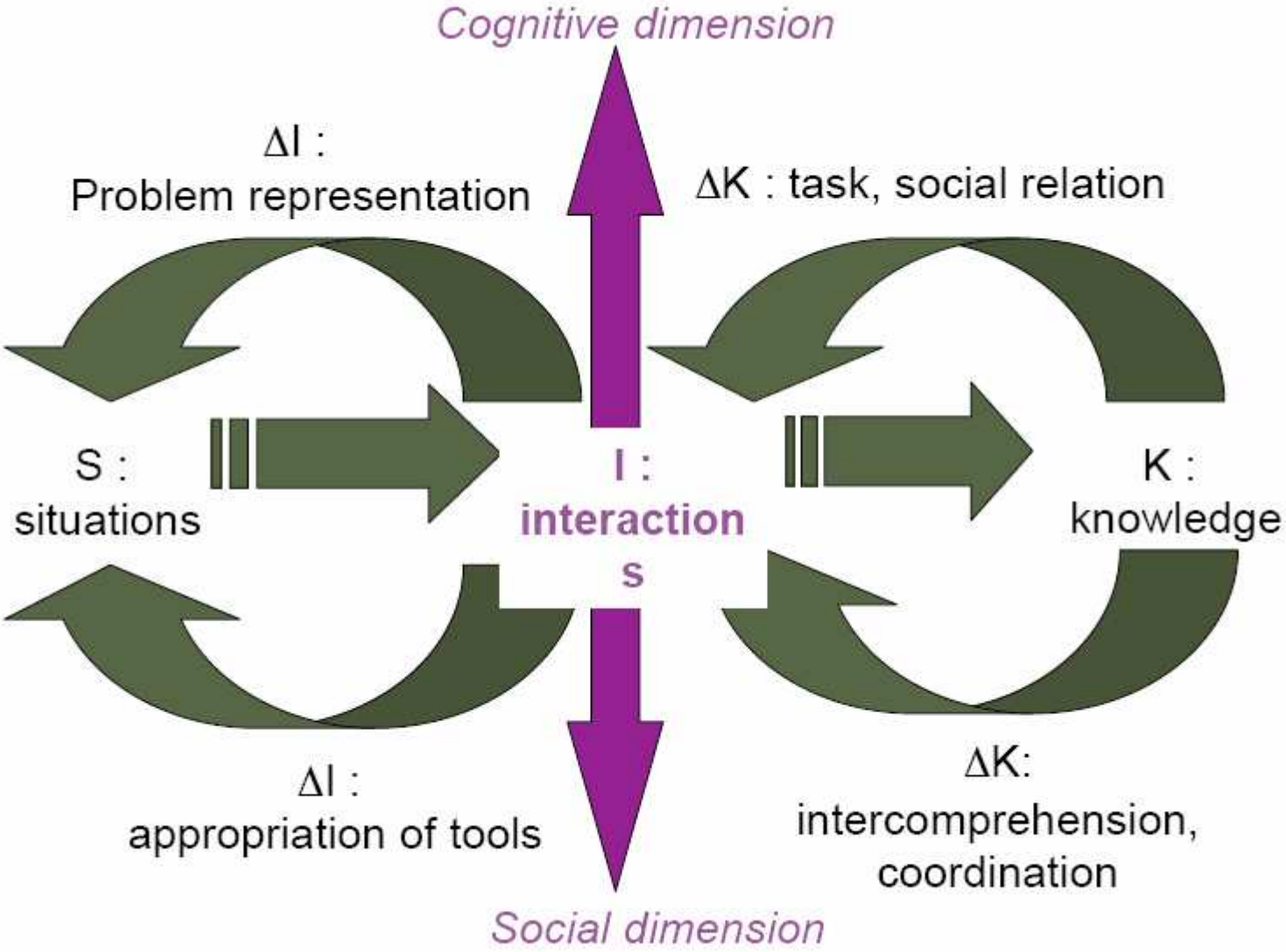


- one of the central challenges of research on CSCL will be the developing of pedagogical models and methodological approaches.
- In order to answer both the cultural and pedagogical challenges, it seems that we also need to explore further the nature of computer supported communication and inquiry itself.
- there has been a change in the research on CSCL to more detailed research on the characteristics of discourse and argumentation.
- Accompanying this process, there is also a need to develop new ways of assessing the learning outcomes in computer supported collaborative learning, because the traditional assessment methods are not necessarily able to show the benefits gained through this kind of learning (e.g. Karlgren).

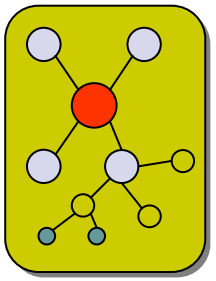
CSCL: space and time

SPACE \ TIME	Synchronous	Asynchronous
Co-present		
Distant	 	 

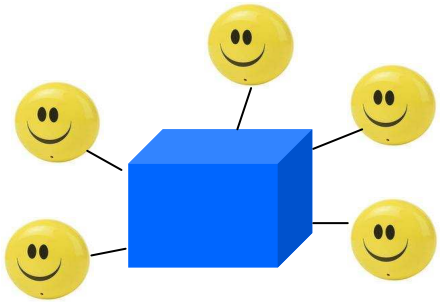
“Constructive Interaction” paradigm

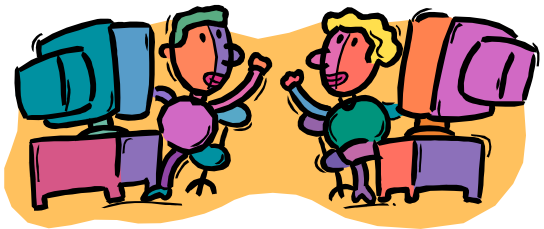


A New Approach

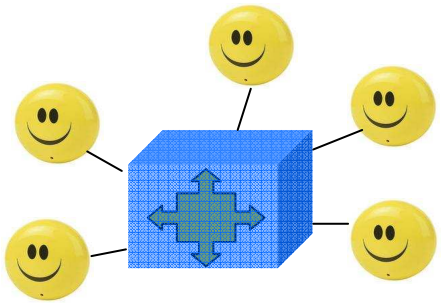


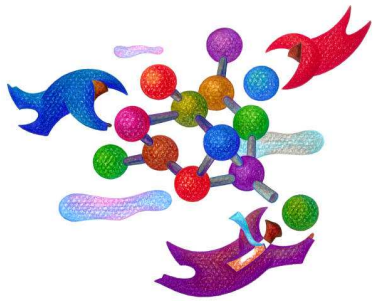
- It's still unclearly about how the group member implement Collaborative Knowledge Building in interaction. (Koschmann)
- We should study and understand categories of interactions and the mechanisms of negotiation to a much greater depth rather than only study collaboration in general . (Dillenburg & Baker)





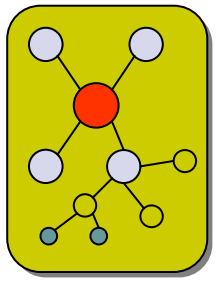
Effect of Collaboration =Post-test — Pre-test

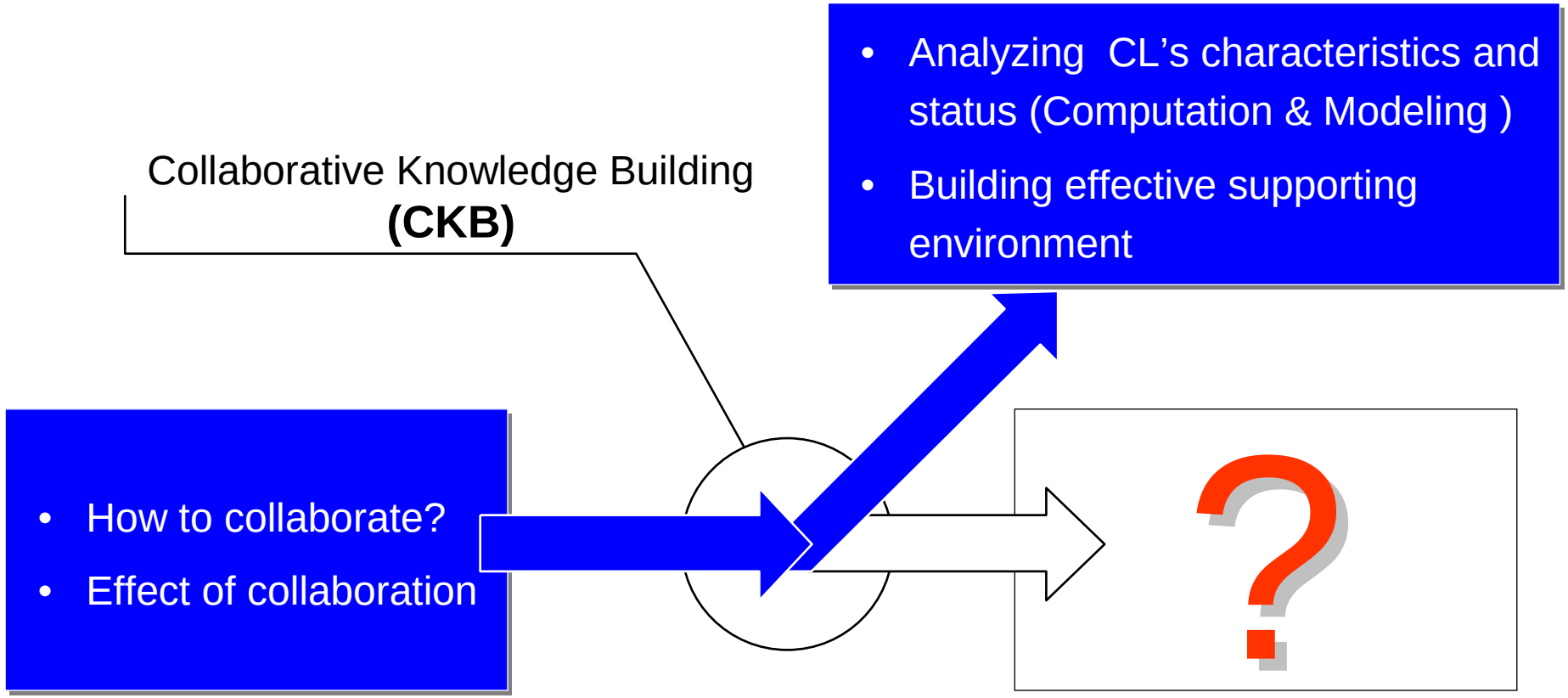




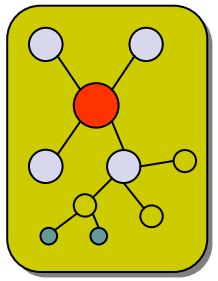
Process of Collaborative Knowledge Building

A New Approach – Content Analysis?



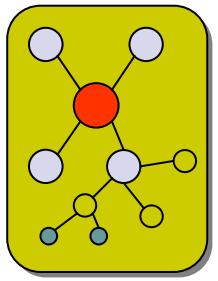


Materials in content analysis



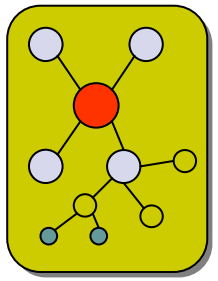
- Coding the **content of documents** (like behavior coding)
  - any technique for making inferences by identifying special characteristics of messages (**written or oral**)
    - **artifacts** of social communications
  - information is condensed (**classified**) and made systematically comparable by applying a **coding scheme**
- any kind of **written document**
  - field notes from participant observation, letters, novels, transcripts of recorded communications (such as T.V shows, interviews, etc.)

The steps in content analysis



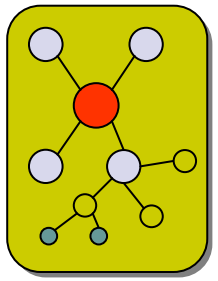
1. fully describe the **phenomenon** to be studies (e.g. portrayal of the elderly in the media)
2. select the **media** that will be used for data
3. derive coding categories
    - **choose categories**, e.g. status of character, physical attractiveness, context, etc.
    - **count** presence or absence of a category
    - place each piece into one of many categories (forced choice)
4. decide on a **sampling strategy** --you can't count it all
5. **train the coders/raters** (***reliability*** is important)
6. **analyze the data** (%'s, compare means and variances?)

Research Perspective



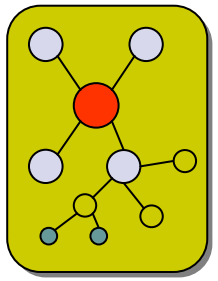
- CL (Collaborative Learning) & CKB (Collaborative Knowledge Building)
  - Interaction in e-Learning Environment is communication among group member mediated by computer and verbal information. Although its form is various, its essential function is Collaborative Knowledge Building, whose essential feature is Collaboration.
  - CKB ,which is interpreted from genetic epistemology ,is the activity that the collaborative community build the shared understanding and form the inter-depended relationship by the interaction among members under the same learning and cultural background.

Research Aim

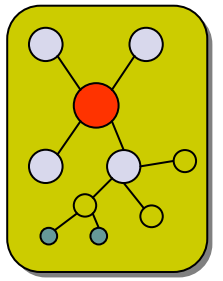


- To find features of CKB and to analyze status of CL
- Explore the approach to extract the above feature from interaction corpus .To design and develop the tools to implement auto analysis or semi-auto analysis of interaction.
- To improve CKB based on the result of interaction analysis. To design and develop CKB Supported Tools.

Analysis Framework: Henri Model

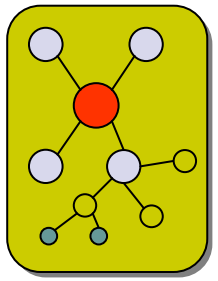


Dimension	Definition	Examples of Indicators
Participative	Compilation of the number of messages or statements transmitted by one person or group	Number of messages Number of statements
Social	Statement or part of statement not related to formal content of subject matter	Self-introduction Verbal support 'I'm feeling great.....!'
Interactive	Chain of connected messages	'In response to Celine.....' 'As we said earlier.....'
Cognitive	Statements exhibiting knowledge and skills relating to learning processes	Asking questions Making inferences Formulating hypotheses
Meta-cognitive	Statements related to general knowledge and skills and showing awareness, self-control, and self -regulation of learning	Commenting on own manner of accomplishing a task Being aware of the emotional context of task completion

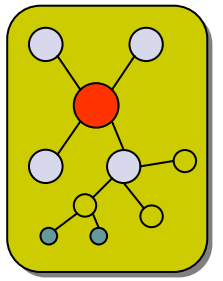


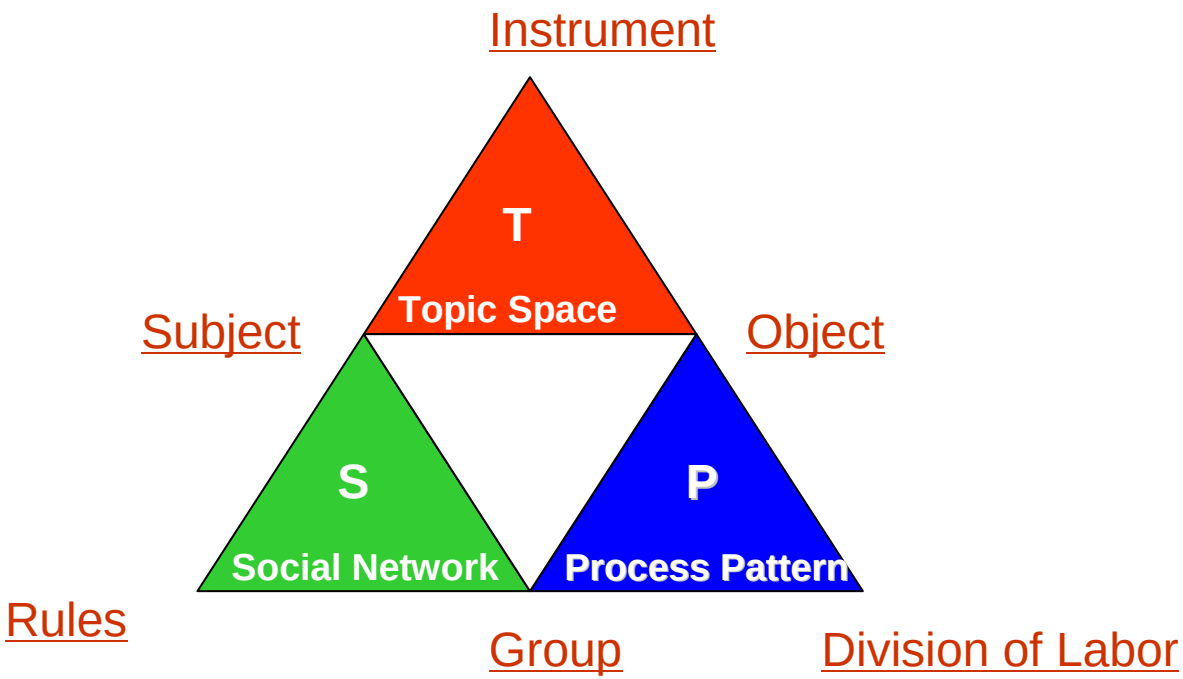
Analytical Category (after Henri, 1992)	Female Student	Male Student	Female Student	Male Student
Social	26%	35%	13%	27%
Interactive	37%	26%	37%	27%
Cognitive Skills (surface)	11%	0%	31%	5%
Cognitive Skills (deep)	26%	39%	13%	32%
Meta-cognitive Knowledge	0%	0%	3%	9%
Meta-cognitive Skills	0%	0%	3%	0%

Phase I	**Sharing/Comparing of Information**	
A	Statements of observation/opinion	6
B	Agreement from one or more participants	4
C	Corroborating examples from one or more participants	6
D	Clarification by asking or answering questions	9
E	Definition, description or identification of a problem	2
Phase II	**Discovery and Exploration of Dissonance or Inconsistency**	
A	Identifying and stating areas of disagreement	6
B	Asking and answering questions to clarify sources of disagreement	2
C	Restating a position and supporting it with evidence	1
Phase III	**Negotiation of Meaning and Co-Construction of Knowledge**	
A	Negotiation or clarification of the meaning of terms	
B	negotiation of the relative weight to be assigned to types of argument	
C	Identification of areas of agreement or overlap among conflicting concepts	
D	Proposal and negotiation of new statements embodying compromise/co-construction	
E	Proposal of integrating or accommodating ideas	
Phase IV	**Testing and Modification of Proposed Synthesis/Co-construction**	
A	Testing synthesis against established/received shared ideas of participants	
B	Testing against existing cognitive schema	
C	Testing against personal experience	
D	Testing against formally collected data	
E	Testing against contradictory testimony in the literature	
Phase V	**Agreement/application of New Co-constructions**	
A	Summarisation of agreements	
B	Application of new knowledge	
C	Metacognitive statements indicating changes of understanding among participants	



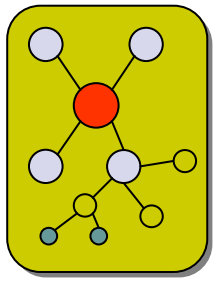
Research framework on CKB Process

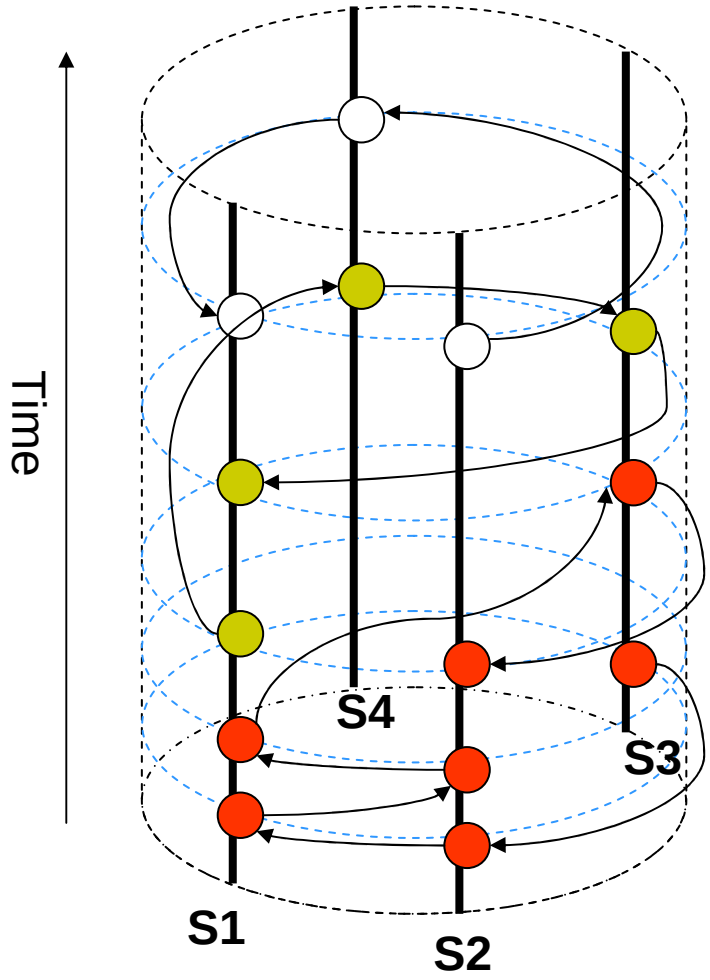




- **Topic Space:** Topic (knowledge) set in process of CKB **.**
- **Social Network:** Relationships among group members .
- **Process Pattern:** The relatively steady path and organizing form of CKB.

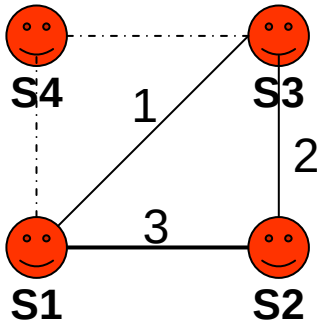
TSP Model



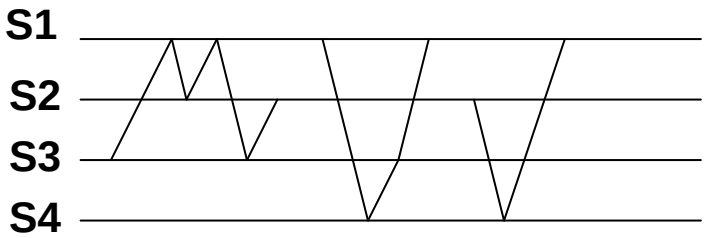


Topic Space：

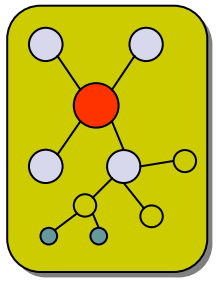
Social Relationship：

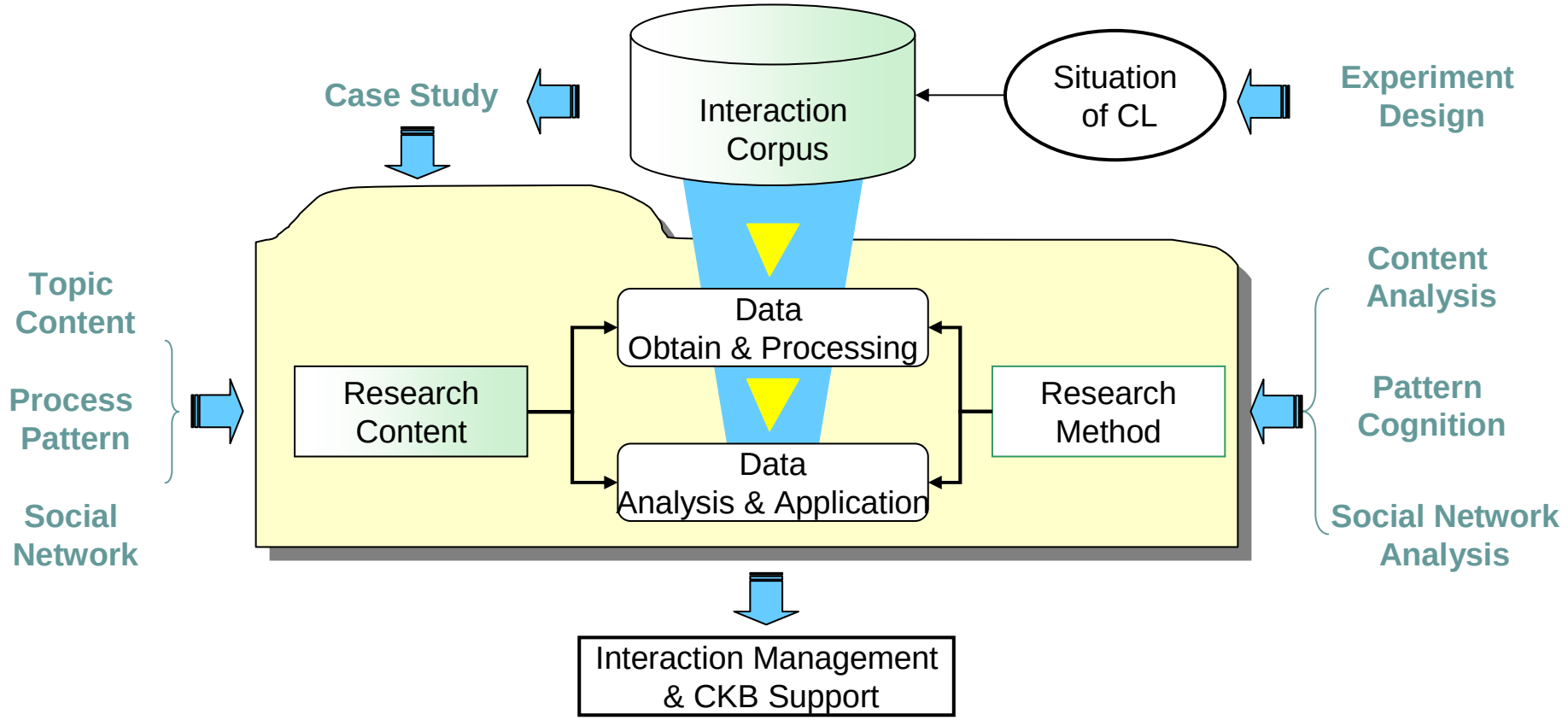


Process Pattern：



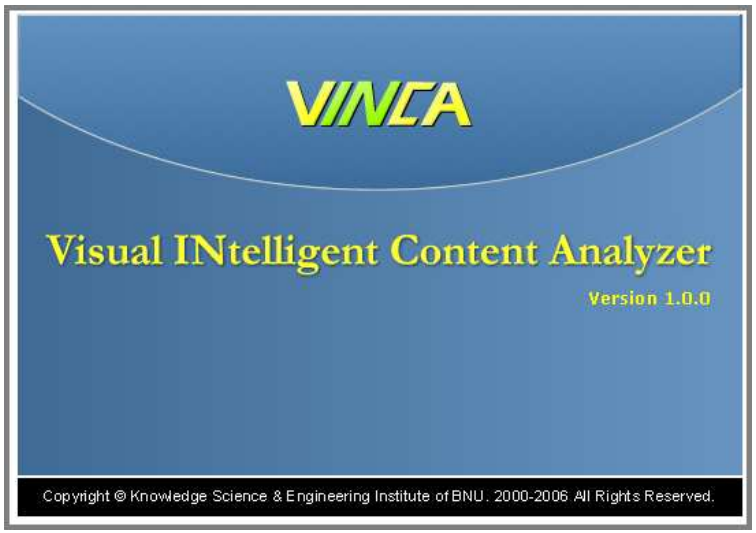
Research Approach



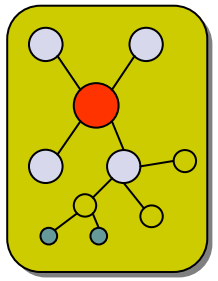


Vinca Introduction

- VINCA stands for *Visual INtelligent Content Analyzer*, which is the content analysis tool jointly developed by CITE, HKU and KSEI, BNU.

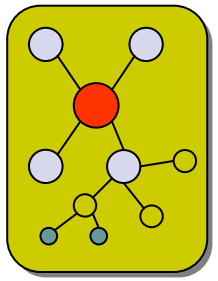


Vinca ‘s Features



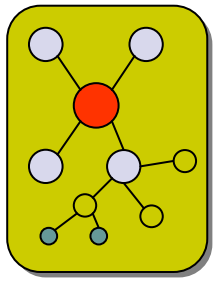
- Learnable Semi-automatic Coding Support
- Analyze text in Chinese
- Utilize Computational Language & Text Mining technologies
- Support assessing for CKB

Vinca's Functions



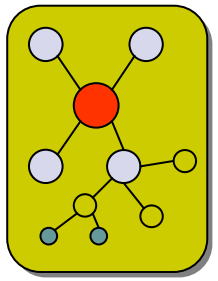
- Data preparation
- Annotation Aids
- Text Analysis
- Data Export for SNA

Data preparation

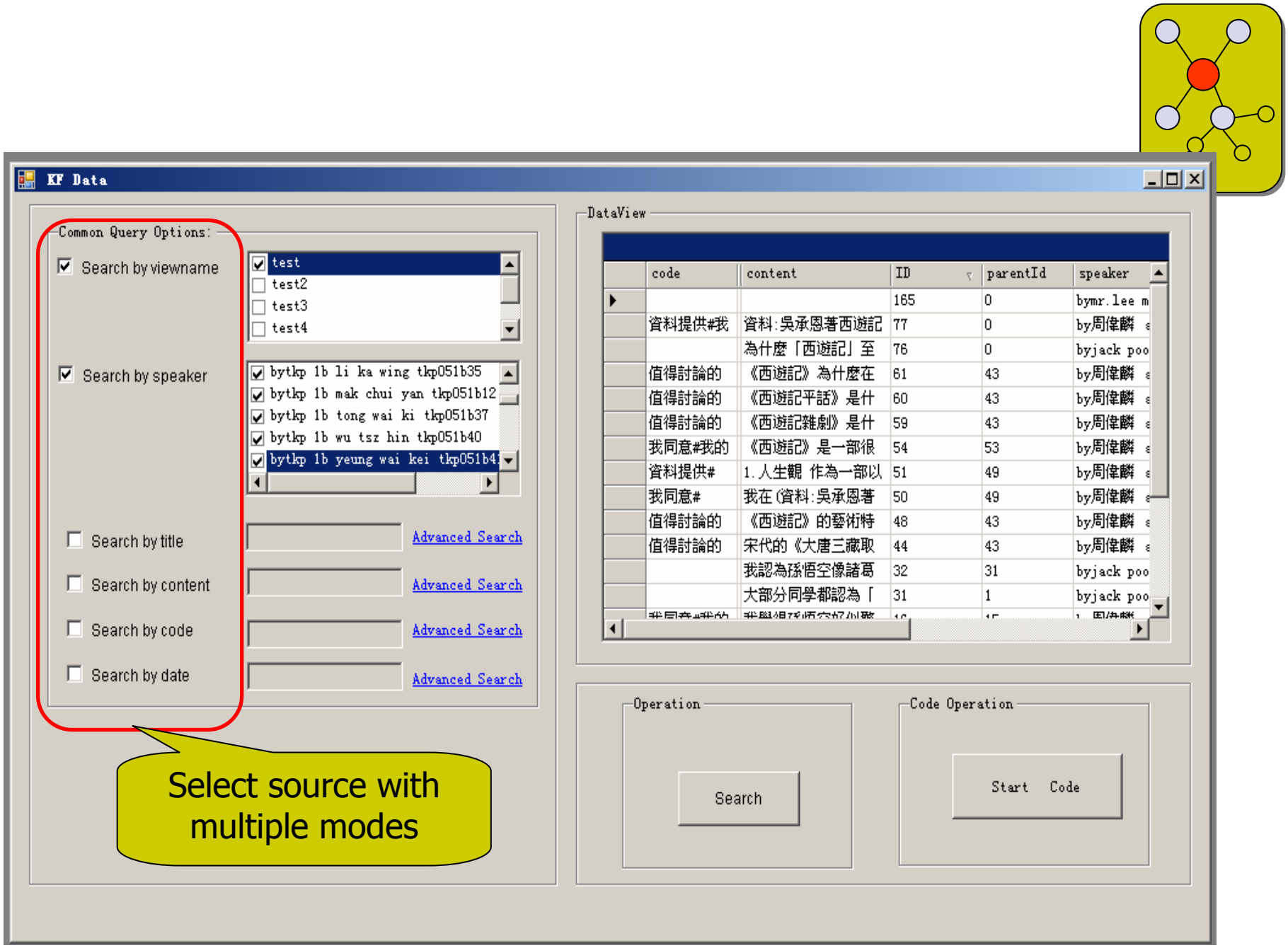


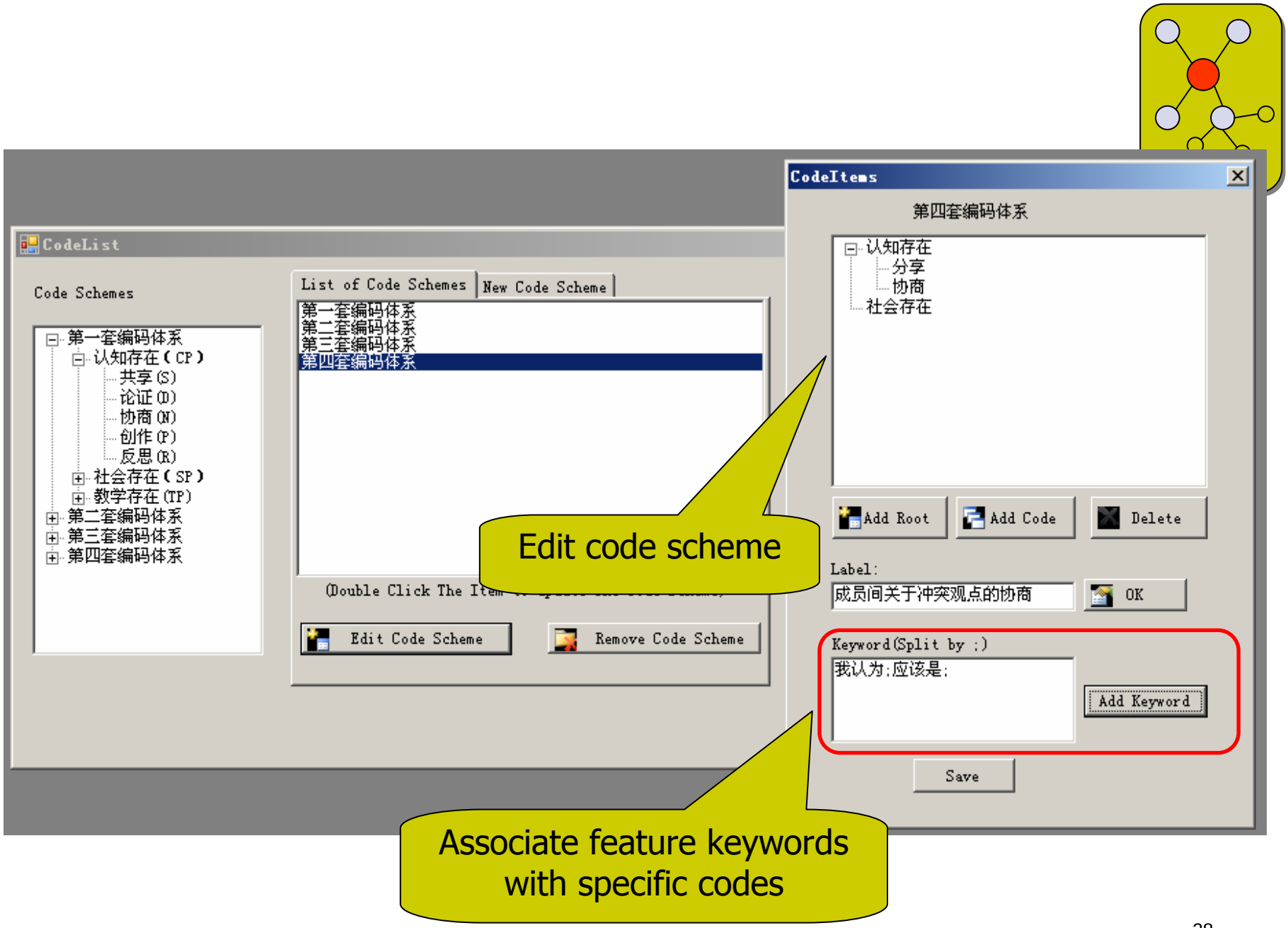
- Data preparation to convert Knowledge Forum discourse in html to database format
  - From Version 3.4
  - From Version 4.5

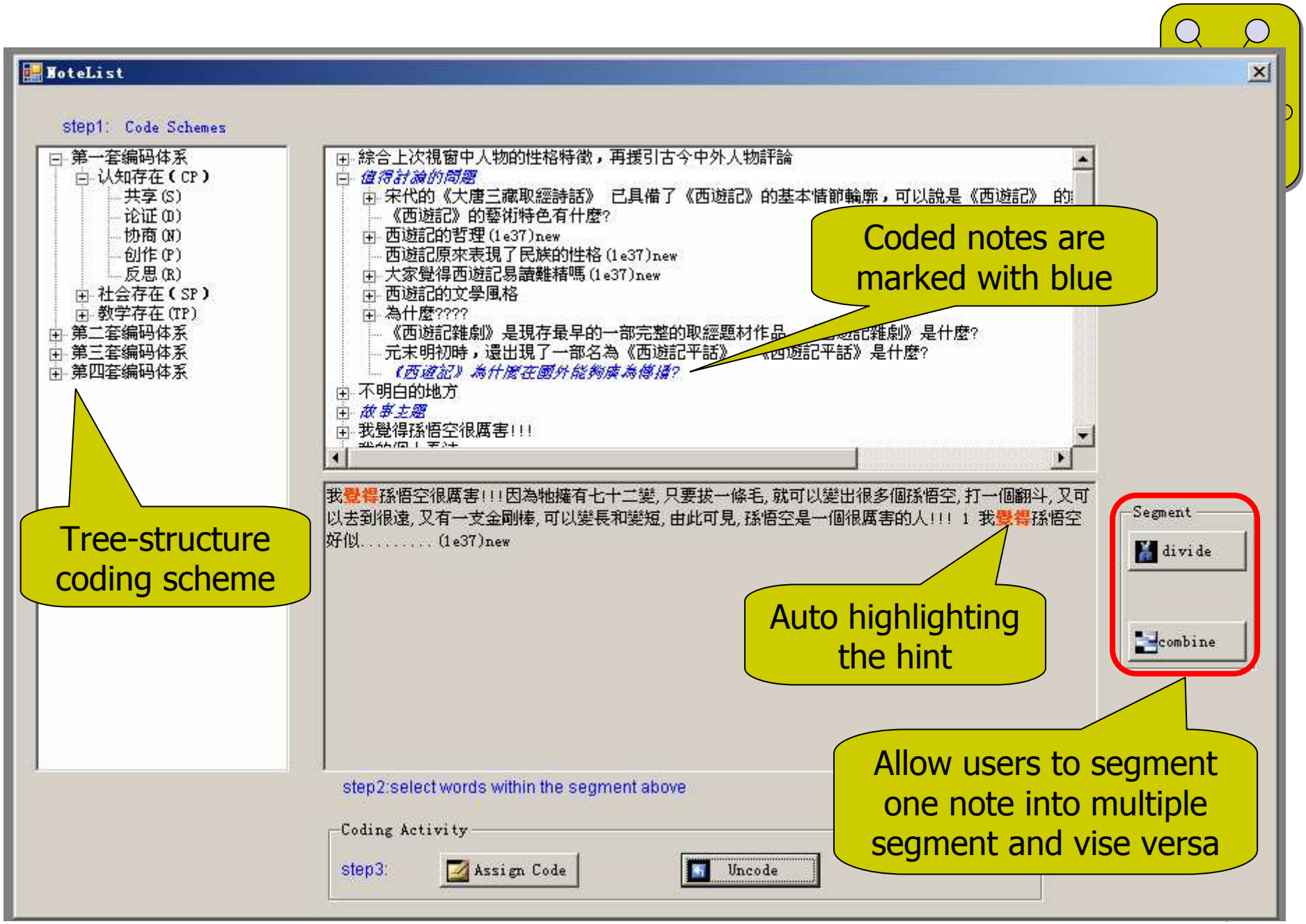
Annotation Aids

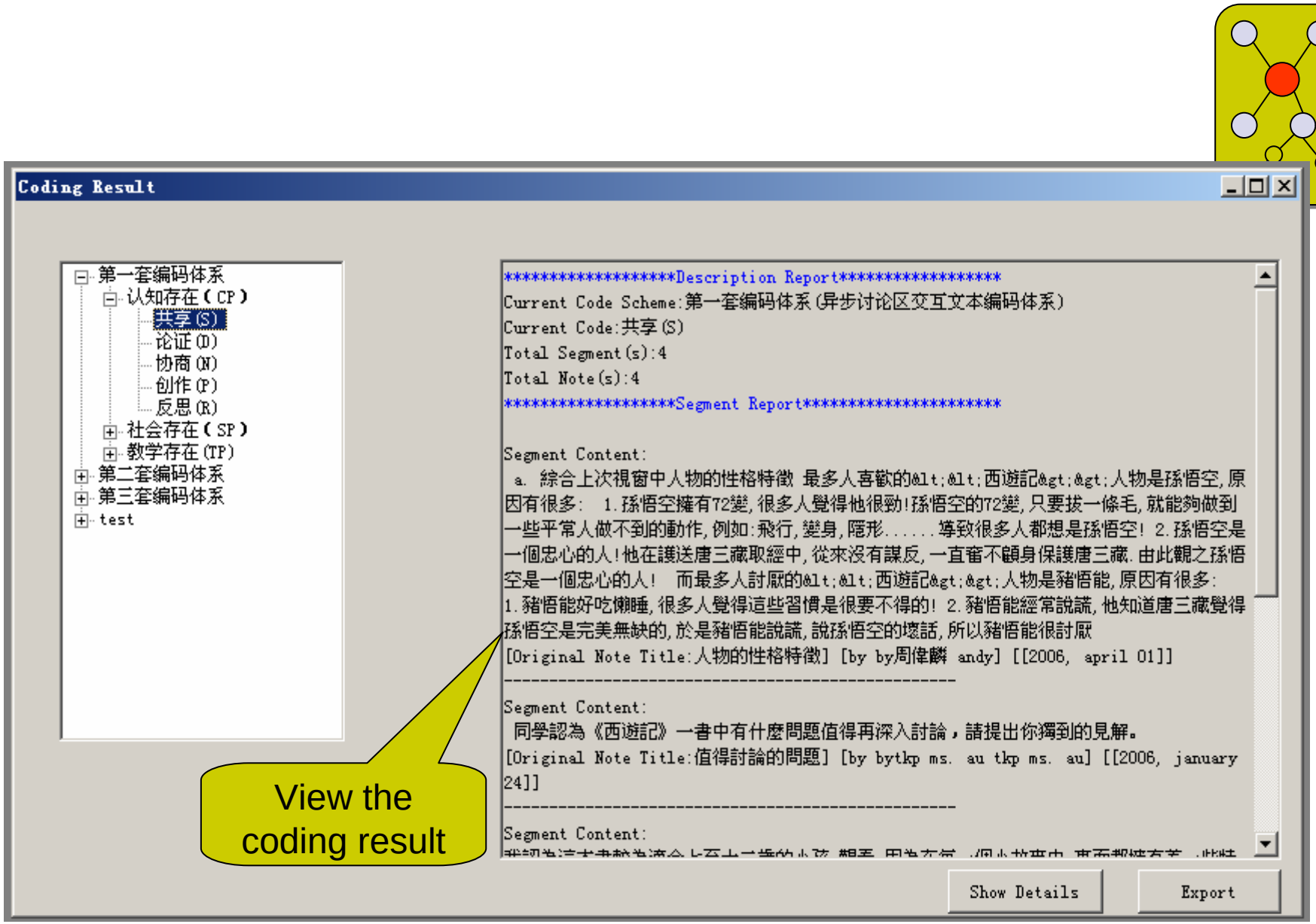


- Edit Coding Scheme
  - New, Modify, Delete
    - Associate feature keywords to specific codes
- Annotation
  - Automatic discover the code hint, highlight it and attach possible codes with confidence probability.
  - Support segment & merge
  - During the process of coding, users are allowed to select the hint to mark the final coding.
- View Coding Result

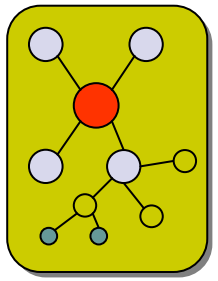

KF Data
Common Query Options:
Search by viewname
test
test2
test3
test4
Search by speaker
bytkp 1b li ka wing tkp051b35
bytkp 1b mak chui yan tkp051b12
bytkp 1b tong wai ki tkp051b37
bytkp 1b wu tsz hin tkp051b40
bytkp 1b yeung wai kei tkp051b4
Search by title
Advanced Search
Search by content
Advanced Search
Search by code
Advanced Search
Search by date
Advanced Search
Select source with multiple modes
DataView
code
content
ID
parentId
speaker
165
0
bymr.lee m
資料提供#我
資料:吳承恩著西遊記
77
0
by周偉麟
為什麼「西遊記」至
76
0
byjack poo
值得討論的
《西遊記》為什麼在
61
43
by周偉麟
值得討論的
《西遊記平話》是什
60
43
by周偉麟
值得討論的
《西遊記雜劇》是什
59
43
by周偉麟
我同意#我的
《西遊記》是一部很
54
53
by周偉麟
資料提供#
1.人生觀 作為一部以
51
49
by周偉麟
我同意#
我在(資料:吳承恩著
50
49
by周偉麟
值得討論的
《西遊記》的藝術特
48
43
by周偉麟
值得討論的
宋代的《大唐三藏取
44
43
by周偉麟
我認為孫悟空像諸葛
32
31
byjack poo
大部分同學都認為「
31
1
byjack poo
Operation
Search
Code Operation
Start Code


CodeList
Code Schemes
第一套编码体系
认知存在(CP)
共享(S)
论证(D)
协商(N)
创作(P)
反思(R)
社会存在(SP)
教学存在(TP)
第二套编码体系
第三套编码体系
第四套编码体系
List of Code Schemes
New Code Scheme
第一套编码体系
第二套编码体系
第三套编码体系
第四套编码体系
Edit Code Scheme
Remove Code Scheme
CodeItems
第四套编码体系
认知存在
分享
协商
社会存在
Add Root
Add Code
Delete
Label:
成员间关于冲突观点的协商
OK
Keyword(Split by ;)
我认为;应该是;
Add Keyword
Save
Edit code scheme
Associate feature keywords with specific codes

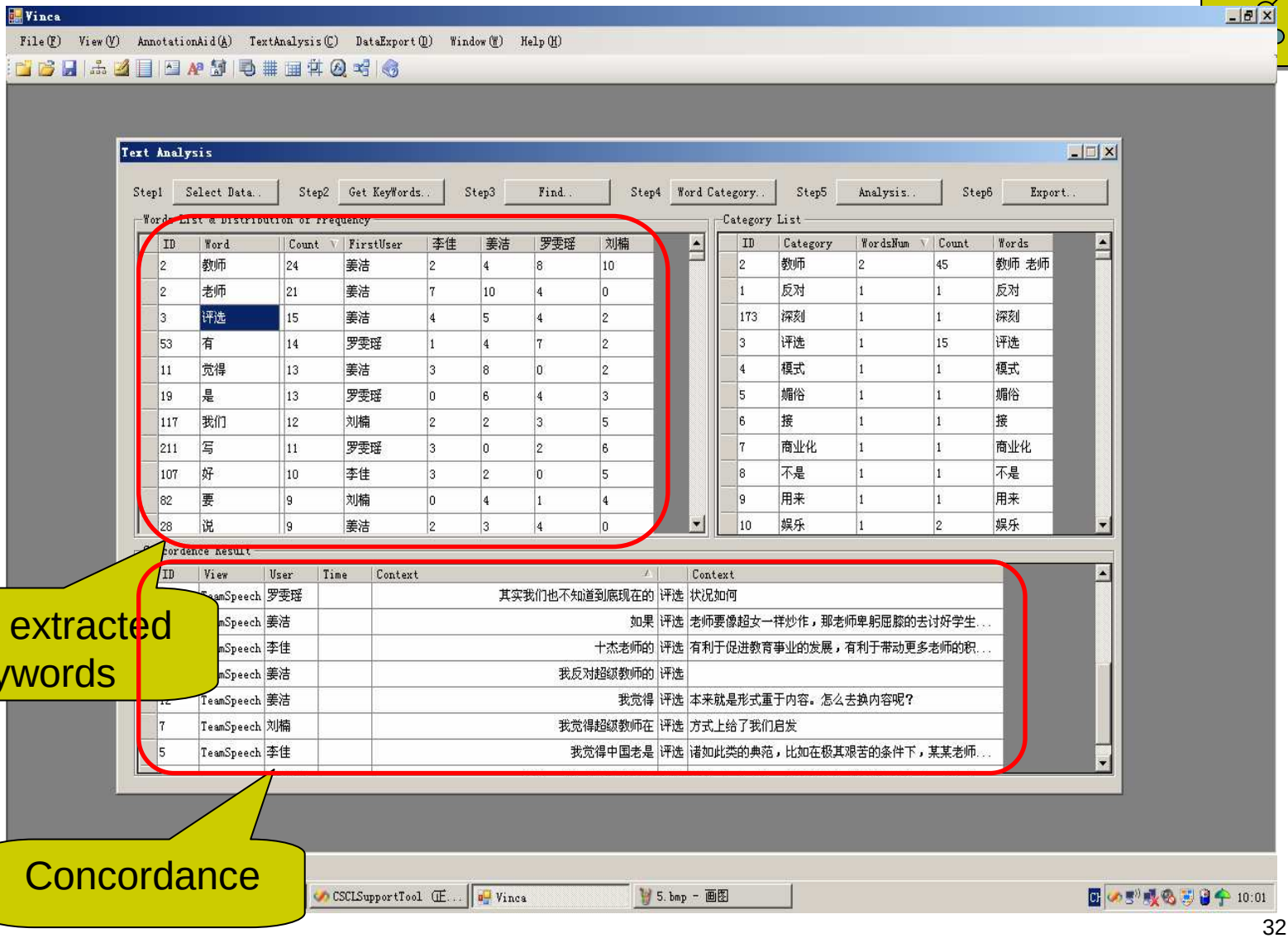

NoteList
step1: Code Schemes
第一套编码体系
认知存在(CP)
共享(S)
论证(D)
协商(N)
创作(P)
反思(R)
社会存在(SP)
教学存在(TP)
第二套编码体系
第三套编码体系
第四套编码体系
綜合上次視窗中人物的性格特徵，再援引古今中外人物評論
值得討論的問題
宋代的《大唐三藏取經詩話》 已具備了《西遊記》的基本情節輪廓，可以說是《西遊記》 的
《西遊記》的藝術特色有什麼?
西遊記的哲理(1e37)new
西遊記原來表現了民族的性格(1e37)new
大家覺得西遊記易讀難精嗎(1e37)new
西遊記的文學風格
為什麼????
《西遊記雜劇》是現存最早的一部完整的取經題材作品
元末明初時，還出現了一部名為《西遊記平話》
《西遊記》為什麼在國外能夠廣為傳播?
不明白的地方
故事主題
我覺得孫悟空很厲害!!!
我覺得孫悟空很厲害!!!因為牠擁有七十二變,只要拔一條毛,就可以變出很多個孫悟空,打一個翻斗,又可以去到很遠,又有一支金剛棒,可以變長和變短,由此可見,孫悟空是一個很厲害的人!!! 1 我覺得孫悟空好似.....(1e37)new
Segment
divide
combine
step2:select words within the segment above
Coding Activity
step3:
Assign Code
Uncode
Coded notes are marked with blue
Tree-structure coding scheme
Auto highlighting the hint
Allow users to segment one note into multiple segment and vise versa

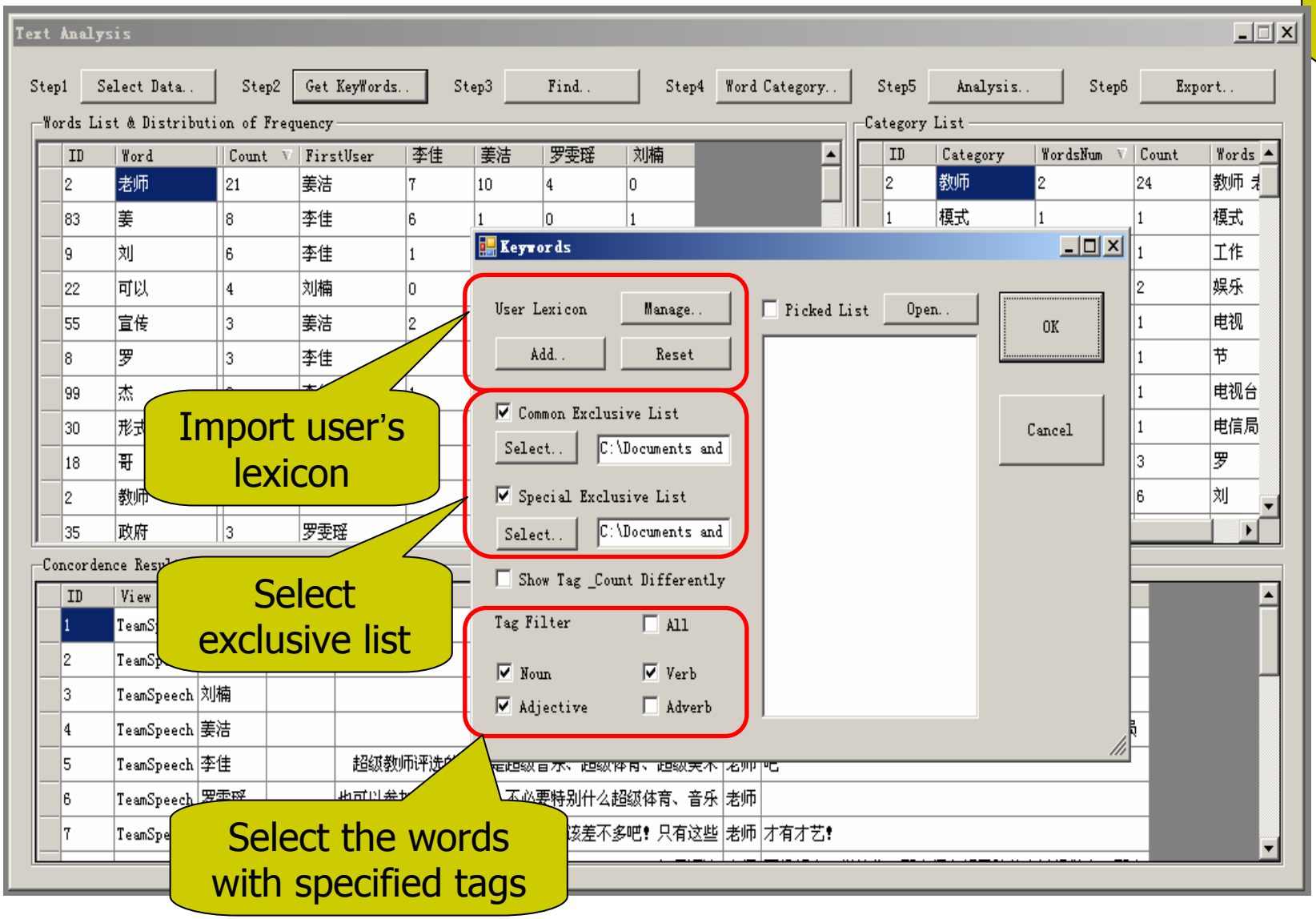

Coding Result
第一套编码体系
认知存在（CP）
共享(S)
论证(D)
协商(N)
创作(P)
反思(R)
社会存在（SP）
教学存在(TP)
第二套编码体系
第三套编码体系
test
**********Description Report**********
Current Code Scheme:第一套编码体系(异步讨论区交互文本编码体系)
Current Code:共享(S)
Total Segment(s):4
Total Note(s):4
**********Segment Report**********
Segment Content:
a. 綜合上次視窗中人物的性格特徵 最多人喜歡的<<西遊記>>人物是孫悟空,原因有很多: 1.孫悟空擁有72變,很多人覺得他很勁!孫悟空的72變,只要拔一條毛,就能夠做到一些平常人做不到的動作,例如:飛行,變身,隱形.....等致很多人都想是孫悟空! 2.孫悟空是一個忠心的人!他在護送唐三藏取經中,從來沒有謀反,一直奮不顧身保護唐三藏.由此觀之孫悟空是一個忠心的人! 而最多人討厭的<<西遊記>>人物是豬悟能,原因有很多:
1.豬悟能好吃懶睡,很多人覺得這些習慣是很要不得的! 2.豬悟能經常說謊,他知道唐三藏覺得孫悟空是完美無缺的,於是豬悟能說謊,說孫悟空的壞話,所以豬悟能很討厭
[Original Note Title:人物的性格特徵] [by by周偉麟 andy] [[2006, april 01]]
Segment Content:
同學認為《西遊記》一書中有什麼問題值得再深入討論，請提出你獨到的見解。
[Original Note Title:值得討論的問題] [by bytkp ms. au tkp ms. au] [[2006, january 24]]
Segment Content:
Show Details
Export
View the coding result

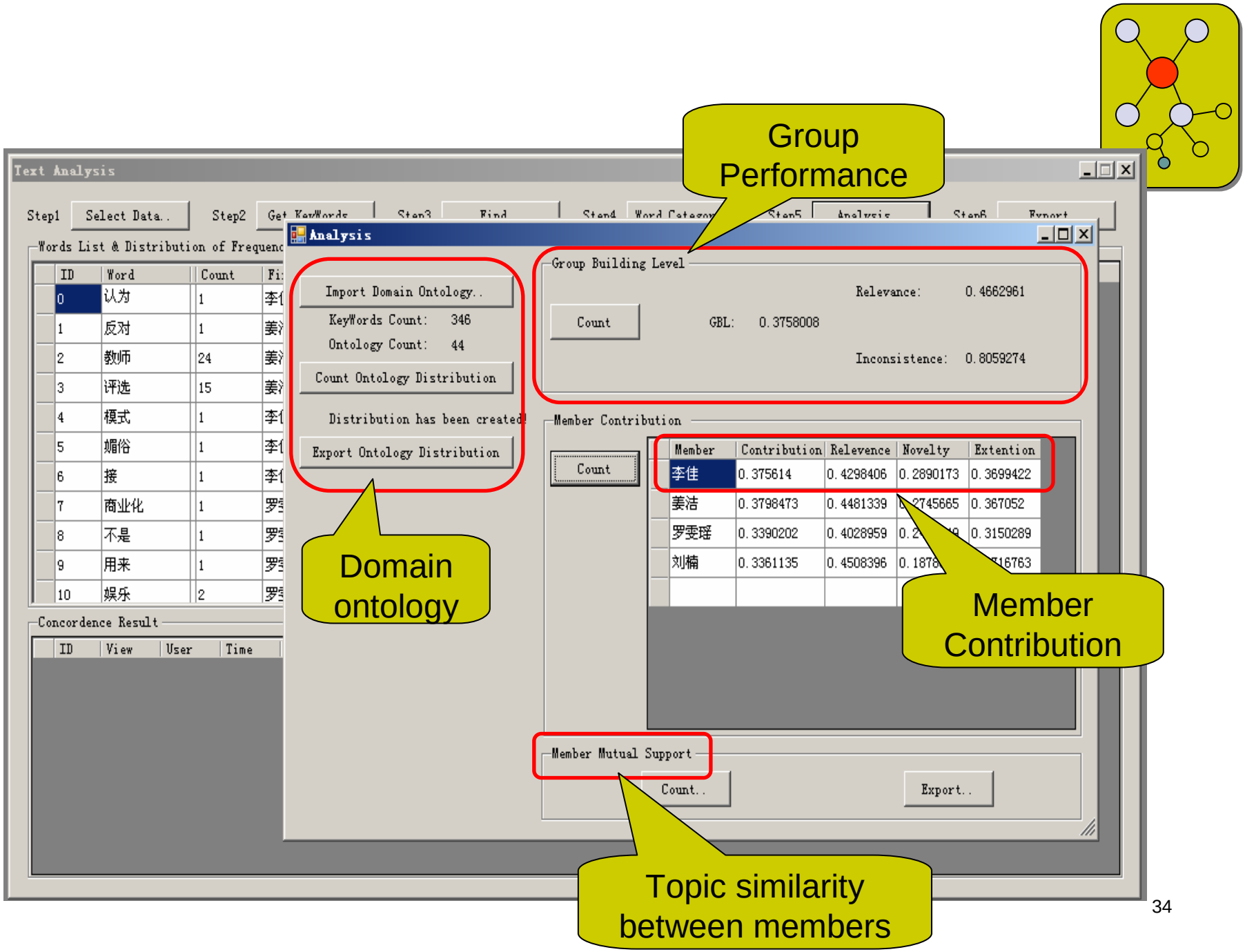
Text Analysis



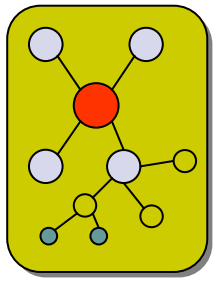
- Keywords retrieval & frequency counting
- Concordance
- Domain ontology-based category analysis
- Text Clustering
- Support assessing for CKB


Vinca
File(F) View(V) AnnotationAid(A) TextAnalysis(C) DataExport(D) Window(W) Help(H)
Text Analysis
Step1 Select Data..
Step2 Get KeyWords..
Step3 Find..
Step4 Word Category..
Step5 Analysis..
Step6 Export..
Words List & Distribution of Frequency
ID Word Count FirstUser 李佳 姜洁 罗雯瑶 刘楠
2 教师 24 姜洁 2 4 8 10
2 老师 21 姜洁 7 10 4 0
3 评选 15 姜洁 4 5 4 2
53 有 14 罗雯瑶 1 4 7 2
11 觉得 13 姜洁 3 8 0 2
19 是 13 罗雯瑶 0 6 4 3
117 我们 12 刘楠 2 2 3 5
211 写 11 罗雯瑶 3 0 2 6
107 好 10 李佳 3 2 0 5
82 要 9 刘楠 0 4 1 4
28 说 9 姜洁 2 3 4 0
Category List
ID Category WordsNum Count Words
2 教师 2 45 教师 老师
1 反对 1 1 反对
173 深刻 1 1 深刻
3 评选 1 15 评选
4 模式 1 1 模式
5 媚俗 1 1 媚俗
6 接 1 1 接
7 商业化 1 1 商业化
8 不是 1 1 不是
9 用来 1 1 用来
10 娱乐 1 2 娱乐
Concordance Result
ID View User Time Context Context
TeamSpeech 罗雯瑶 其实我们也不知道到底现在的 评选 状况如何
TeamSpeech 姜洁 如果 评选 老师要像超女一样炒作，那老师卑躬屈膝的去讨好学生...
TeamSpeech 李佳 十杰老师的 评选 有利于促进教育事业的发展，有利于带动更多老师的积...
TeamSpeech 姜洁 我反对超级教师的 评选
TeamSpeech 姜洁 我觉得 评选 本来就是形式重于内容。怎么去换内容呢?
7 TeamSpeech 刘楠 我觉得超级教师在 评选 方式上给了我们启发
5 TeamSpeech 李佳 我觉得中国老是 评选 诸如此类的典范，比如在极其艰苦的条件下，某某老师...
List of extracted keywords
Concordance
CSCLSupportTool (正...
Vinca
5.bmp - 画图
10:01


Text Analysis
Step1
Select Data..
Step2
Get KeyWords..
Step3
Find..
Step4
Word Category..
Step5
Analysis..
Step6
Export..
Words List & Distribution of Frequency
Category List
Keywords
User Lexicon
Manage..
Add..
Reset
Picked List
Open..
OK
Cancel
Common Exclusive List
Select..
C:\Documents and
Special Exclusive List
Select..
C:\Documents and
Show Tag _Count Differently
Tag Filter
All
Noun
Verb
Adjective
Adverb
Import user's lexicon
Select exclusive list
Select the words with specified tags


Group Performance
Text Analysis
Step1
Select Data..
Step2
Words List & Distribution of Frequenc
ID
Word
Count
0
认为
1
1
反对
1
2
教师
24
3
评选
15
4
模式
1
5
媚俗
1
6
接
1
7
商业化
1
8
不是
1
9
用来
1
10
娱乐
2
Concordence Result
ID
View
User
Time
Analysis
Import Domain Ontology..
KeyWords Count: 346
Ontology Count: 44
Count Ontology Distribution
Distribution has been created!
Export Ontology Distribution
Group Building Level
Count
GBL: 0.3758008
Relevance: 0.4662961
Inconsistence: 0.8059274
Member Contribution
Count
Member
Contribution
Relevence
Novelty
Extention
李佳
0.375614
0.4298406
0.2890173
0.3699422
姜洁
0.3798473
0.4481339
0.367052
罗雯瑶
0.3390202
0.4028959
0.3150289
刘楠
0.3361135
0.4508396
Member Mutual Support
Count..
Export..
Domain ontology
Member Contribution
Topic similarity between members

Data Export for SNA



- Export KF Data
- Export Relation Matrix
- Export Coding Result
- Export Coding Matrix
- Export Coding Frequency

Thanks

Welcome Questions and Comments